Support Project for Data Fusion Computation: Current status and future prospects

Shin'ya Nakano[1,2]*

[1]* *The Institute of Statistical Mathematics, ROIS, Midoricho 10-3, Tachikawa, Tokyo, 190-8562, Japan*
[2] *School of Multidisciplinary Science, SOKENDAI, Hayama, Kanagawa, 240-0193, Japan*
Email: shiny@ism.ac.jp

Summary. Support Project for Data Fusion Computation (SPDFC) is a project aiming at providing our knowledge of novel statistical techniques for simulation researchers. Some statistical approaches are useful for enhancing the effectiveness of numerical simulation. One example is data assimilation which estimates a scenario of temporal evolution by incorporating a sequence of the observational data into a numerical simulation model. Another example is a statistical emulator which imitates a simulation model by a statistical model. This paper briefly explains these statistical techniques useful for simulation researches.

Keywords. data assimilation, statistical emulator, numerical simulation.

1. Introduction

Numerical simulation is widely used in variety of fields such as weather prediction, fishery prediction, aeroplane designing, building designing, and so on. Recent simulation models can compute the evolution of a system with high temporal and spatial resolution. Some of the simulation models achieves so high accuracy that a simulation result is hard to discriminate from a real phenomenon.

Numerical simulation just reproduces a scenario of temporal evolution of a system under given conditions. Thus, when conducting numerical simulation, it is crucial to give appropriate conditions such as initial conditions and boundary conditions. Statistical approaches are useful for appropriately setting the conditions to be given for simulation. However, simulation researchers are not necessarily familiar with statistical approaches which would be useful for enhancing the effectiveness of their simulation models. In order to promote application researches of statistical techniques for enhancing effectiveness of numerical simulation, Research Institute of Information and Systems (ROIS) has launched Support Project for Data Fusion Computation (SPDFC) as one of working groups under Joint Support-Center for Data Science Research in ROIS. Under this project, we provide the knowledge of novel statistical techniques for simulation researchers at request. We also give seminars and hands-on for a wide range of simulation researchers. In addition, we are developing statistical software for implementing the statistical techniques to various simulation programmes.

This paper explains key statistical techniques for enhancing simulation researches, which are promoted under this project. The activities for promoting such statistical techniques are also introduced.

2. Statistics for simulation

There are two important statistical techniques for enhancing numerical simulation: data assimilation and statistical emulation. In this section, we explain each of the two techniques.

2.1 Data assimilation

In order to accurately reproduce temporal evolution of a real phenomenon by a simulation, it is essential to give an initial condition and boundary condition which well correspond to the real situation. In most cases, the observational data are limited and we can observe only a small fraction of the entire initial state and boundary state. However, a time series of the observational

data enables us to constrain the initial and boundary states more effectively. In addition, a simulation model can also be used as a effective constraints about the temporal evolution of the system because the simulation model is composed on the basis of our knowledge on the physical laws governing the system.

Data assimilation [1] estimates a scenario of temporal evolution by incorporating a sequence of the observational data into a numerical simulation model. Data assimilation not only utilise the information of the observational data taken at various times but also exploit the simulation model as a constraint for the estimation. Data assimilation has been developed as a fundamental technique for numerical weather prediction. Numerical weather prediction uses an atmospheric simulation model which describes physical processes in the atmosphere. The atmospheric system can be regarded as a deterministic but chaotic system. In order to successfully predict a future weather, it is crucial to accurately estimate the initial condition. Even in other various fields, it would be essential to obtain a good initial condition to predict evolution of a system. The applications of data assimilation techniques are widely expanding.

2.2 Statistical emulator

In engineering applications, numerical simulation is conducted for finding an optimal value of tuneable design parameters. When designing a complex system, it is difficult to guess the effect of the change of such tuneable parameters. In such cases, numerical simulation is helpful for evaluating the effect of tuneable parameters. However, in order to find an optimal value of tuneable parameters for a complex system, simulation must be run many times until the optimal value is found. It tends to take much time to obtain one scenario by running a simulation model of a complex system. Thus, it is sometimes practically impossible to run such a simulation model repeatedly until the optimal value is found.

In order to avoid to repeatedly run a computationally expensive simulation model, a statistical model imitating a simulation model is considered [2-3]. Such a statistical model is called a 'statistical emulator' or an 'emulator'. A statistical emulator is derived by analysing the relationship between the input and the output using a statistical technique. The statistical emulator is a potentially useful tool for designing a complex system.

3. Concluding remarks

We are now conducting several collaborative researches of applications of data assimilation techniques and statistical emulation techniques. These collaborative researches cover a variety of fields including material science and social science. In addition, we are developing software in order to facilitate simulation researchers in various fields who want to conduct data assimilation. A data assimilation problem can be formulated by a common framework called a state space model. Thus, the software is designed to be applicable to problems in various fields as far as interfaces for a simulation model and data access are prepared in a pre-defined manner. Data assimilation techniques tends to requires high computational cost. We are thus developing parallelised software which can be run on a supercomputer.

References

1. Kalnay, E., Atmospheric modeling, data assimilation and predictability. *Cambridge University Press, Cambridge, UK,* 2003
2. Kennedy, M. C., O'Hagan, A., Bayesian calibration of computer models. *J. Roy. Statist. Soc. Ser. B,* 63, 425-464, 2001
3. Rougier, J., Efficient emulators for multivariate deterministic functions. *J. Comp. Graph. Statist.,* 17, 827-843, 2008